Bothalia, 10, 3: 437–450

Principal Components Analysis of Acacia burkei and A. nigrescens in Natal

by

J. H. Ross* and J. W. Morris*

Abstract

Four principal component analyses were carried out to study the perplexing relations within *Acacia burkei* Benth. and between it and *A. nigrescens Oliv.* Sampling methods are described in detail. Ten morphological parameters were noted from 163 plants of 21 populations. The results confirmed conclusions of earlier non-multivariate studies. The two species can be distinguished on the basis of the ten parameters and it is of doubtful value to recognize infraspecific categories within *A. burkei* as the variation within the species is continuous. The technique of principal components analysis was most useful in this study.

Introduction

Acacia burkei Benth. and *A. nigrescens* Oliv. form part of a complex of closely related species which are taxonomically most perplexing. Within this complex the degree of pubescence of the calyx is the character of prime importance in distinguishing two main groups. In their typical forms *A. nigrescens* and *A. burkei* are readily distinguishable: the former with its large leaflets and glabrous calyces and the latter with smaller leaflets and pubescent calyces. However, there are numerous plants with leaflets intermediate in shape and in size between those of *A. nigrescens* and those of *A. burkei*. Leaflet size varies considerably and an entire range from those the size of *A. burkei* to those the size of *A. nigrescens* may be found on a single plant. However, as these plants have pubescent calyces their relationship seems to be with *A. burkei* rather than with *A. nigrescens*.

This range of morphological variation within *A. nigrescens* and within *A. burkei* has been considered in some detail (Ross 1968a, 1968b). It had been customary to distinguish loosely between "small leaflet" *A. burkei* and "big leaflet" *A. burkei*, the former typically having leaflets less than 3 mm wide and the latter leaflets more than 3 mm wide. However, it was found (Ross 1968b) that the characters typifying "small leaflet" and "big leaflet" *A. burkei* were not necessarily correlated but varied independently, certain combinations of characters being commoner than others. Thus, although specimens at either extreme of the range of morphological variation could be readily sorted into two groups, there remained numerous specimens that could not be referred to either group with certainty. Consequently no infraspecific categories were recognized within *A. burkei*.

An examination of the means of the morphological parameters (see below) obtained for *A. nigrescens* (Ross, 1968a) and for *A. burkei* (Ross, 1968b) indicated that rachilla length, number of pinna pairs, number of leaflet pairs, leaflet length and leaflet width provided discontinuities between the two species. When the extremes of the morphological parameters were examined, however,

* Botanical Research Institute.

these differences were not so readily apparent. Consequently it was decided to subject the morphological parameters to a principal components analysis in an attempt either to confirm or contradict earlier findings that *A. nigrescens* and *A. burkei* are quite readily separated, and that it is of doubtful value to recognize infraspecific categories within such an inherently variable species as *A. burkei*.

Kendall (1957) and Seal (1964) described principal components analysis in detail and one introduction to the subject, in a taxonomic setting, is given by Jeffers (1965). The results of many taxonomic applications have been published recently but as this is one of the first in South Africa the method is described in detail. In the context of this paper the object of the technique is to extract a set of components from the populations x parameters matrix which account for as much as possible of the parameter variation between the *Acacia* populations and which are mathematically independent of one another.

Sampling Technique

To assess the morphological variation within and among trees, and within and among populations, some statistical procedure was essential. The application of statistical methods brought with it the need for reliable, yet practical sampling techniques. The average herbarium collection is unsuitable, consisting often of isolated specimens selected as being "typical", either of a single plant, or of a population, or of aberrants sufficiently atypical to have attracted attention.

The prime requirement for a statistical study is that samples be representative. This proved difficult since populations were not always clearly defined and often occupied rugged terrain. Individual plants because of their large, woody, much-branched growth form and abundant foliage presented yet other sampling problems. All such problems had to be met by employing techniques that yielded representative samples, yet were essentially practicable.

Twenty leaves, twenty pods and twenty inflorescences from each plant were regarded as a satisfactory number for a sample. To obtain such samples from individual plants, a sampling method devised for and tested out on *A. robusta* Burch. (Gordon-Gray, 1965) was employed.

The distal one to two feet of not less than ten branches representative of the crown of a plant were collected.[1] In no instance were coppice shoots included since preliminary work showed that the leaves of such shoots differ, either in size or in pubescence. The branches collected from any one plant constituted a sample.

Each sample was treated separately. All mature leaves were stripped from the branches, heaped together and thoroughly mixed. Immature leaves were ignored. From this heap twenty leaves were taken by an operator with eyes closed. The same procedure was followed to obtain a sample of twenty pods and twenty inflorescences.

As many populations as possible of each species, which occur scattered through Natal (almost entirely in Zululand), were visited and sampled (see Fig. 1). Most populations visited covered large areas. Because of the rugged terrain, plants growing on, or near, the roadside were sampled. Availability alone governed the haphazard intervals at which plants were sampled. As far

1. It was appreciated at the outset that a truly random sampling method, such as 'Randomised Branch Sampling' (Jessen, 1955), was not practicable in this study. Consequently the word 'random' has been omitted throughout, lest its use infringe mathematical requirements. In all sampling procedures followed, however, care was exercised to ensure that samples were representative and without bias.

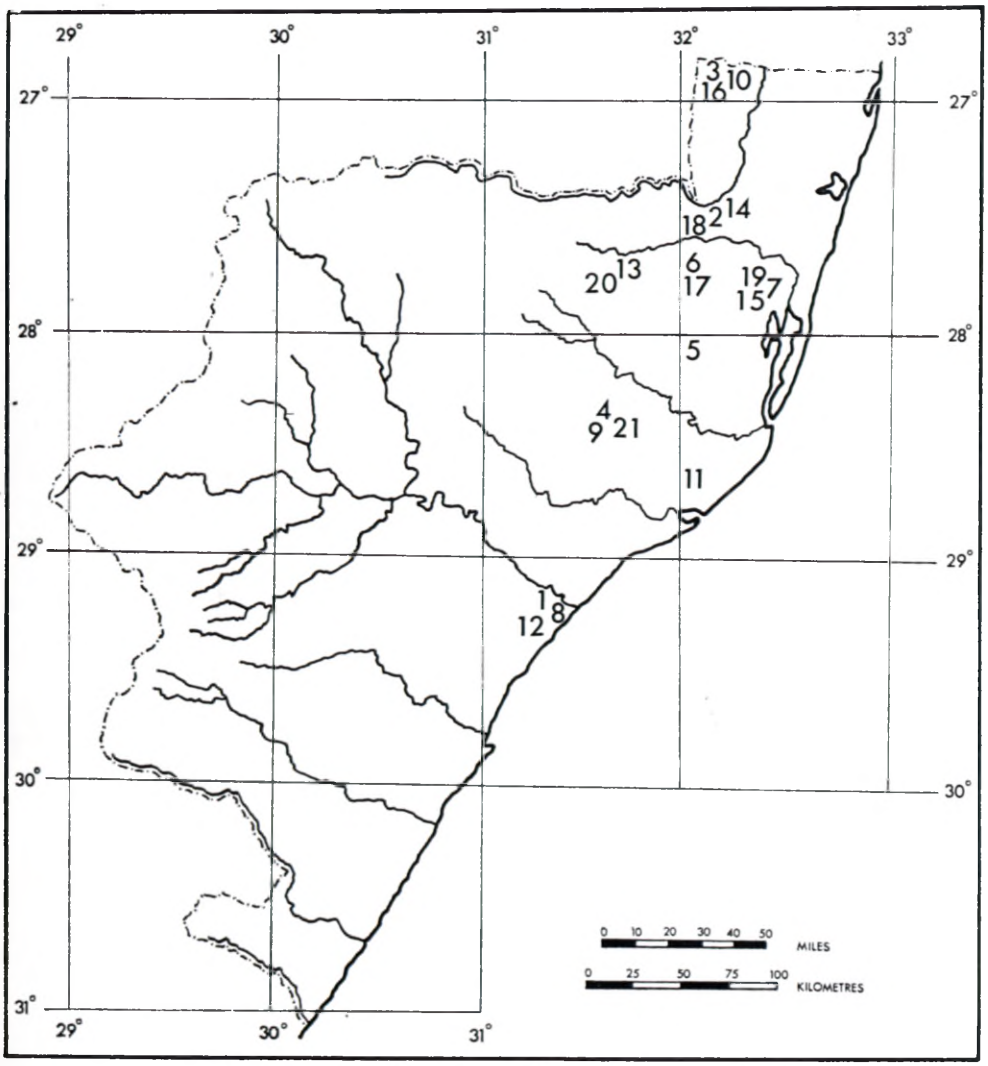

FIG. 1. — Localities of sampled populations in Natal.

as possible, ten plants were sampled from each population, but where populations were small fewer were sampled. A rough sketch map of the population was made on which the positions of the sampled plants were plotted. Each plant sampled was identified by means of a numeral painted on the bark. This was important since it was usually necessary to re-visit the plants as few had flowers and ripe pods contemporaneously.

The morphological parameters for each of the twenty leaves and pods which constituted the sample from each tree were:-

1. Petiole length (mm).
2. Rachis length (mm).
3. Leaf length (mm).
4. Rachilla length (mm) (the length of the right-hand member of the rachilla pair central on the leaf, abaxial surface uppermost).

5. Number of pinna pairs (mean).
6. Number of leaflet pairs (mean).
7. Leaflet length (mm) (length of the leaflet midway along the right-hand member of the rachilla pair central on the leaf, abaxial surface uppermost).
8. Leaflet width (mm) (as for leaflet length).
9. Pod length (mm).
10. Pod width (mm).

Means for each tree, referred to below as tree means, and for each population, referred to as population means, were calculated. Parameter means for the 15 populations of *A. burkei* (1-15) and the six populations of *A. nigrescens* (16-21) are given in Table 1.

TABLE 1. — Population means for each morphological parameter. The parameters are enumerated in the text and the localities of the populations are indicated in Fig. 1.

Population Number	Parameter									
	1	2	3	4	5	6	7	8	9	10
1	24.0	42.7	66.7	22.7	3.0	4.0	11.4	5.2	95.2	14.7
2	20.8	38.6	59.4	20.6	2.0	2.0	11.1	6.5	101.1	19.1
3	18.5	53.6	72.1	28.6	4.4	4.9	13.1	6.2	108.2	18.0
4	17.8	46.8	64.6	23.4	6.4	9.7	7.6	2.7	87.9	15.7
5	17.1	46.1	63.2	26.1	6.3	9.4	7.9	3.1	131.8	17.2
6	16.0	47.0	63.0	24.1	5.6	6.2	8.3	3.8	117.3	21.2
7	15.8	38.0	53.8	19.4	4.5	4.1	8.4	4.4	81.4	20.2
8	15.6	45.0	60.6	25.8	7.8	11.2	5.2	2.1	73.1	16.5
9	15.3	36.0	51.0	20.8	4.3	4.0	10.3	5.9	72.0	18.6
10	14.6	47.7	62.3	25.3	7.6	10.9	6.6	2.6	101.2	16.2
11	13.2	34.6	47.8	18.0	8.0	12.2	4.6	1.7	75.4	18.1
12	12.8	28.4	41.2	19.2	5.6	8.8	5.5	2.3	78.3	17.9
13	12.8	31.8	44.6	20.4	4.0	3.0	9.5	5.1	84.3	18.7
14	11.0	34.3	45.3	19.5	7.1	10.7	5.1	1.7	71.5	16.9
15	10.2	31.4	41.6	20.0	8.6	11.1	4.1	1.4	104.1	18.3
16	19.9	42.3	62.2	9.5	2.9	1.0	22.9	18.7	109.4	20.8
17	17.1	40.8	57.9	8.0	3.1	1.0	21.3	17.3	105.1	15.2
18	14.8	38.2	53.0	6.6	3.1	1.0	20.9	15.6	106.0	17.1
19	14.4	37.2	51.6	8.3	3.2	1.0	20.7	17.1	106.7	17.9
20	14.4	29.9	44.3	10.8	3.0	1.0	22.7	17.1	108.3	17.3
21	13.7	28.8	42.5	7.5	3.0	1.0	18.0	14.6	115.0	16.6

Data Analysis

Four principal component analyses were performed on the available data. Firstly, population means for both *A. burkei* and *A. nigrescens* were used and secondly, tree means for both species were used. The third and fourth analyses were carried out on, respectively, population means and tree means for *A. burkei* alone. Thus the raw data matrix (population x parameters) for the first analysis contained 21 population means, the second 163 tree means, the third 15 population means and the fourth 118 tree means. In each analysis all ten morphological parameters were used. The raw data for the first and third analysis are given in Table 1. Shortage of space precludes inclusion of the raw data for the second and fourth analysis but it is available from the authors on request.

For each analysis the first step was the computation of correlation co-efficients between each parameter and each other one over all population or tree means, resulting in a symmetrical 10 x 10 matrix. The principal components were extracted from this matrix.

An eigenvalue and eigenvector are associated with each principal component. The value indicates the proportion of the total variation accounted for by the component and thus the "importance" of the component, and the vector gives the weighting of each parameter. Components are extracted in descending order of eigenvalues, hence the name principal components. The vector is scaled so that the highest value is unity. In practice it has been found that parameters having weightings of over 0.7 and under —0.7 are important, the importance being proportional to the absolute value.

Two-dimensional scatter diagrams were constructed from the analyses. The position of a population along an axis is found by summing the products of the eigenvector and parameter vector for the population.

Results and Discussion

First Analysis

Eigenvalues and eigenvectors resulting from the first analysis are given in Tables 2 and 3, respectively. Inspection of Table 2 shows that almost half the variation within the correlation matrix is extracted by the first component, that over 90 per cent is extracted by the first four components and virtually all is extracted by the first six components. Further discussion will be limited to the first three components which account for 88 per cent of the variation.

Table 2. — Eigenvalues of the first seven components extracted by the first analysis.

Component	Eigenvalue	Percentage of variability	
		Component	Cumulative
1	4.846	48.467	48.467
2	2.937	29.379	77.847
3	1.057	10.577	88.424
4	0.557	5.575	94.000
5	0.358	3.583	97.583
6	0.184	1.840	99.424
7	0.044	0.449	99.873

TABLE 3. — Eigenvectors of the first three components extracted by the first analysis.

Parameter Number	Eigenvectors corresponding to component: 1	2	3
1	0.295	*0.848*	—0.833
2	—0.176	*0.926*	0.149
3	—0.030	*1.000*	0.052
4	*—0.836*	0.493	—0.045
5	*—0.930*	—0.161	0.274
6	*—0.968*	—0.037	0.141
7	*1.000*	0.008	—0.012
8	*0.992*	—0.102	0.042
9	0.699	0.383	0.282
10	0.206	—0.074	*1.000*

The morphological parameters contributing most to the first component's variation are 4, 5, 6, 7 and 8 (values given in italics in Table 3). The first three parameters and the last two are positively correlated between themselves, but the two groups are negatively correlated. Parameter 9 also has a high weighting on the first component. Parameters 1, 2 and 3 contribute most to the second component and parameter 10 is the only important one on the third component.

Positions of the populations along the first and second and first and third components are given in Fig. 2. A clear discontinuity between *A. burkei* and *A. nigrescens* is shown along the first component. There is also a discontinuity

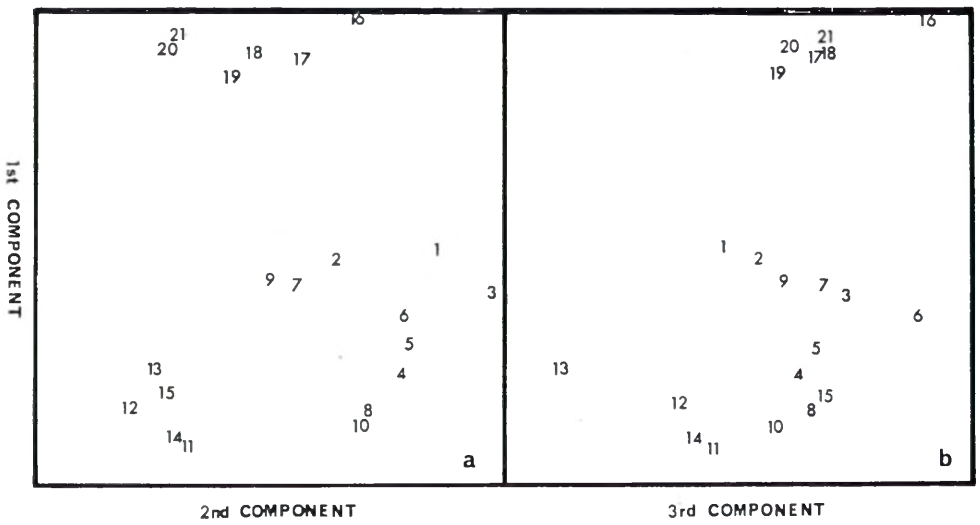

FIG. 2. — Positions of **A. burkei** (1-15) and **A. nigrescens** (16-21) populations plotted against the first and second components (a) and the first and third components (b) of the first analysis.

on the second component within *A. burkei*. However, on the third component there are no distinct discontinuities within *A. burkei* although population 13 is somewhat separate from the remaining populations. Within the *A. nigrescens* populations a cluster is formed by all the populations except 16 which is somewhat separate, particularly along the third component. Very little variation is evident within *A. nigrescens* along the first component.

Figure 2 indicates a definite distinction between *A. burkei* and *A. nigrescens* based on the sampled populations and on the morphological parameters used, and suggests that *A. burkei* is a more variable species than *A. nigrescens*. As almost three times more *A. burkei* than *A. nigrescens* populations were sampled, it is not possible to conclude with certainty that the former is the more variable species, but a trend which supports findings of previous, non-multivariate studies (Ross 1968a, 1968b) is evident. The reason for the greater variation within *A. burkei* has not been studied.

Parameters responsible for the separation of *A. burkei* and *A. nigrescens* are those mentioned above, with high absolute values within the first eigenvector. Likewise, parameters responsible for the spread amongst *A. burkei* populations along the second and third components are those with high absolute values within the second and third eigenvectors respectively. The values for three parameters which have high absolute values within the first eigenvector are plotted in Figure 3

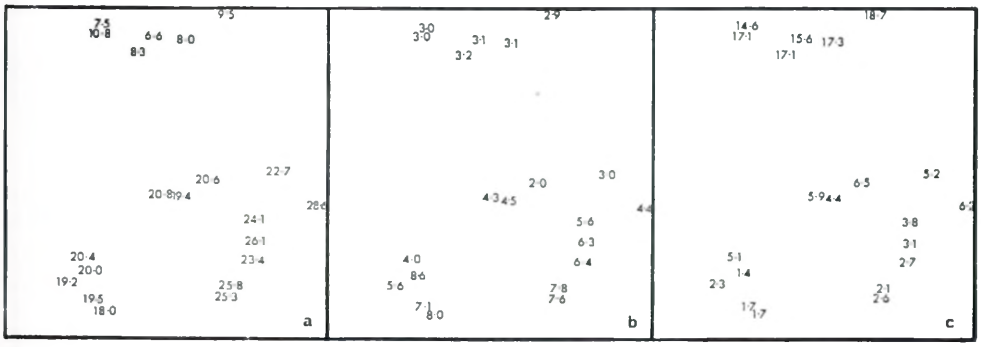

FIG. 3. — (a) Rachilla length (parameter 4), (b) number of leaflet pairs (parameter 6) and (c) leaflet width (parameter 8) of populations plotted against the first and second components of the first analysis.

against the first and second components of Figure 2a. The clear difference between the two species with respect to these parameters can be seen. It will be noticed that better gradients along the first component are shown by number of leaflet pairs and leaflet width than by rachilla length. As rachilla length has an eigenvector value of only 0.636 a very good fit is not expected. The good gradient along the second component is, however, expected as rachilla length has the highest eigenvector value on this component.

In Figure 3 the first two parameters are positively correlated because in both cases the higher values are found amongst the *A. nigrescens* populations. As the higher values for leaflet width are found amongst the *A. burkei* populations, leaflet width is negatively correlated with rachilla length and number of leaflet pairs.

The positive and negative correlations discussed here and earlier had been discovered before the multivariate analysis was undertaken. The negative correlation of parameters is because the longer leaves have relatively fewer pinna pairs and, similarly, long rachillae have relatively fewer pairs of larger leaflets. Conversely, short rachillae carry a larger number of smaller leaflets. The agreement between what was known and the results of the analysis add to one's confidence in the technique.

Table 3 reveals that petiole length, rachis length and leaf length are the most important characters on the 2nd component and that pod width is the most important character on the 3rd component. As these characters mainly affect the distribution of *A. burkei* populations within the ordination, they will be discussed later where analyses without the presence of *A. nigrescens* populations are presented. The first two analyses were undertaken to study the relationship between the two species and not within each.

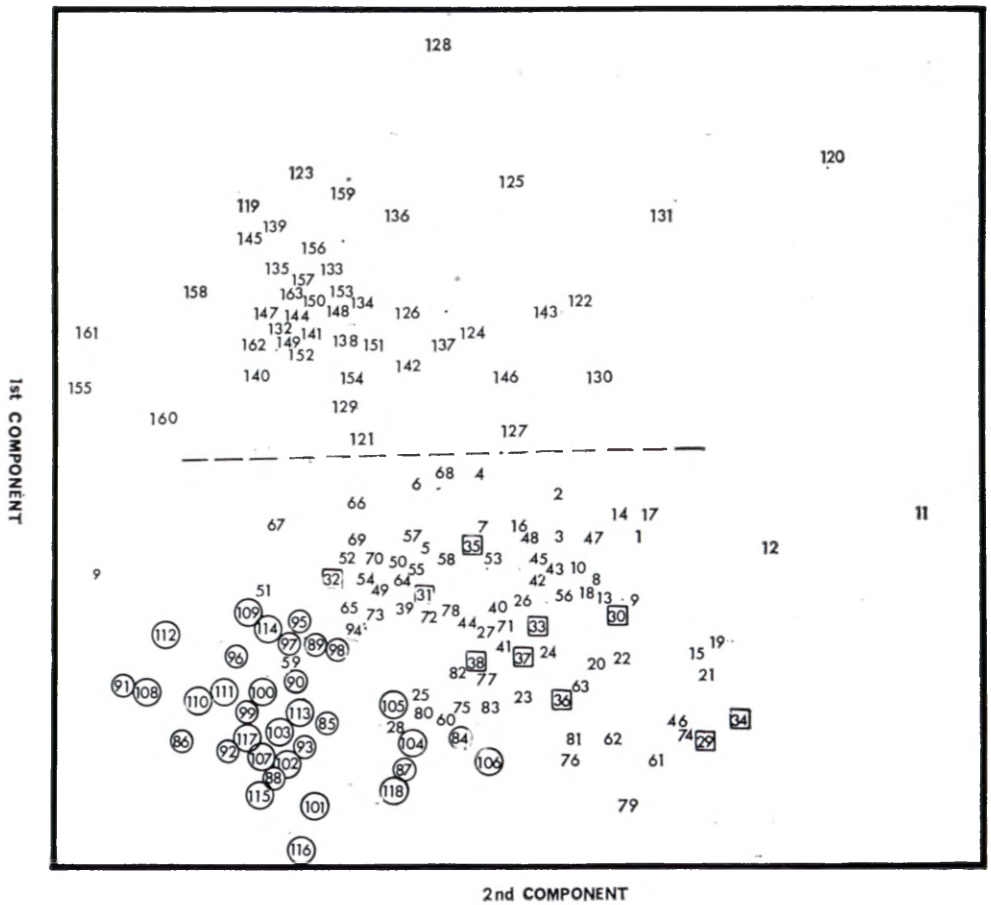

FIG. 4. — Positions of **A. burkei** (1-118) and **A. nigrescens** (119-163) tree means plotted against the first and second components of the second analysis. Means for populations 11-15 are circled and means of population 5 are boxed. A dotted line separates **A. burkei** from **A. nigrescens** populations.

Second Analysis

Tree means for *A. burkei* and *A. nigrescens* are plotted against the 1st and 2nd components in Figure 4. Although there is not a clear discontinuity, the two species are completely separated by the first component.

It is known that *A. burkei* and *A. nigrescens* are distinct species and that they are readily distinguished on the degree of pubescence of the calyx. This character, being of the presence/absence type, was not mixed with the other characters which are approximately continuous. The object of using the data, even though this taxonomically significant character had been omitted, was to establish whether or not the two species could still be separated by multivariate analysis. Certain trees of *A. nigrescens* (e.g. 121 and 127) are similar to certain trees of *A. burkei* (e.g. 4, 6, 68) with regard to the 10 characters sampled but the usual clear distinction between the species is indicated in Figure 4 by the two distinct clusters formed along the first component.

As in Figure 2, there is a tendency for the *A. burkei* trees to be more spread along the 2nd component than are the *A. nigrescens* trees. Populations 11—15, represented in Figure 4 by trees numbered 84—118 (circled), are again in proximity although in this instance there is no discontinuity between these trees and the remainder as in Figure 2. It was appreciated at the outset that population means were of limited value but they give a useful summary of the situation. Comparison of Figure 2 with Figure 4 shows how erroneous a picture can be obtained from the use of population means alone. Furthermore, in Figure 4, where the means of each tree were used, it is seen that there is considerable variation within each population. For example, population 5 of Figure 2 is represented by trees numbered 29—38 in Figure 4.

Third Analysis

For the third analysis population means for *A. burkei* alone were used. Inspection of the eigenvalues showed that over 54 per cent of the variability within the correlation matrix was extracted by the 1st component and over 90

TABLE 4. — Eigenvectors of the first three components extracted by the third analysis.

Parameter Number	Eigenvectors corresponding to component: 1	2	3
1	*0.945*	0.000	—0.141
2	*0.728*	*0.962*	0.057
3	*0.893*	*0.724*	—0.008
4	0.636	*1.000*	—0.063
5	*—0.859*	*0.791*	0.093
6	*—0.780*	*0.907*	—0.073
7	*1.000*	—0.406	—0.030
8	*0.892*	—0.698	0.087
9	0.683	0.580	0.410
10	—0.164	—0.183	*1.000*

per cent by the first three components together. Eigenvectors for the first three components are given in Table 4 and positions of populations plotted against the first and second and first and third components are given in Figure 5.

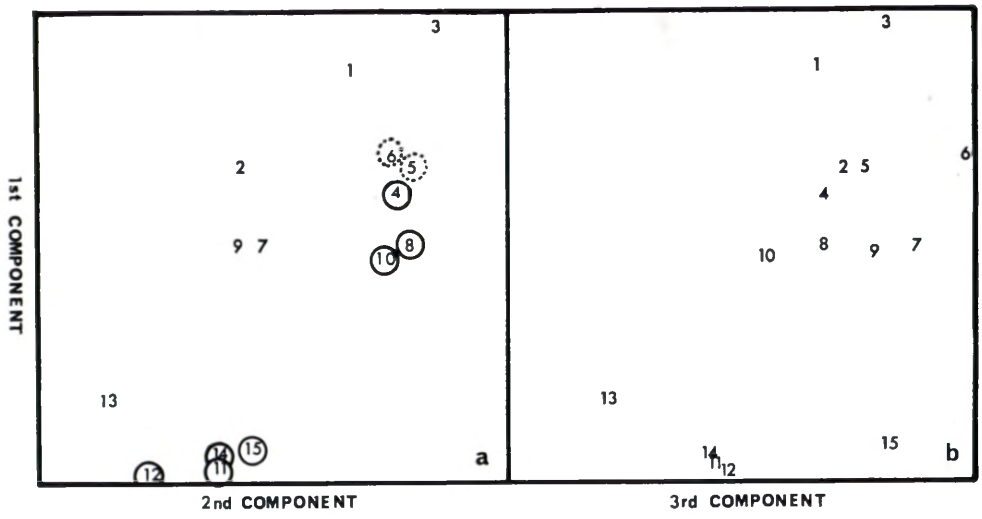

FIG. 5. — Positions of **A. burkei** populations plotted against the first and second components (a) and the first and third components (b) of the third analysis. **A. burkei** "small" populations are circled and mixed "big" and "small" populations are marked by a dotted circle.

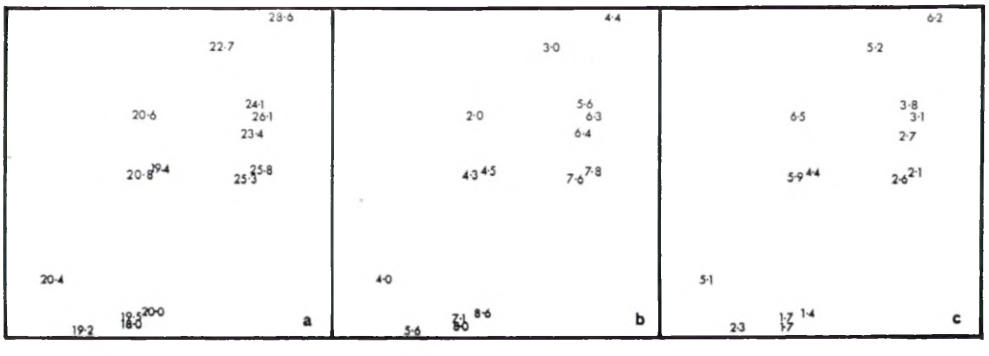

FIG. 6. — (a) Rachilla length (parameter 4), (b) number of leaflet pairs (parameter 6) and (c) leaflet width (parameter 8) of populations plotted against the first and second components of the third analysis.

With the exception of pod width all parameters have relatively high weightings along the first component, while parameters 2 through 6 have high weightings along the second component and pod width is the only important parameter along the third component. Many high weightings are often found on the first component of an analysis and can usually be attributed to overall size differences. However, the first component does not separate "big leaflet" and "small leaflet" populations as one would then expect. Instead, "big" and "small" populations occur scattered along the first axis. This is also shown in Figure 7 (see below).

The second component spreads the populations in such a way that "big" and "small" populations can be separated by a diagonal line extending from between populations 12 and 13 to between 1 and 4. This is, however, the axis of maximum variation along the first two components. This means that the "big" to "small" difference is secondary to another, more important, gradient which separates populations 11 to 15 from the rest.

In Figure 6 the three morphological characters used in Figure 3 were plotted against the first and second components of the third analysis. Similar positive and negative correlations as in Figure 3 are shown. There is an indistinct gradient along the first component in leaflet width and number of pinna pairs. Leaflet width is, however, the character on which "big" plants are separated from "small", once again suggesting that there is some other character, or characters, which are more important than leaflet width in drawing out the populations and splitting populations 11 to 15 from the rest. Table 4 reveals that petiole length, rachis length, leaf length, number of pinna pairs, number of leaflet pairs, leaflet length and leaflet width are all important in creating variation between populations. All of these characters contribute either positively or negatively to the split between populations.

Fourth Analysis

The eigenvalues and eigenvectors of the fourth analysis are very similar to those of the third. The similarity is to be expected as the data for the third analysis are derived directly from those of the fourth. Slightly less variability

TABLE 5. — Eigenvectors of the first three components extracted by the fourth analysis.

Parameter Number	Eigenvectors corresponding to component: 1	2	3
1	*0.896*	0.287	—0.072
2	0.690	*1.000*	0.145
3	*0.835*	*0.871*	0.088
4	0.544	*0.941*	0.128
5	*—0.882*	0.691	0.206
6	*—0.809*	*0.883*	0.097
7	*1.000*	—0.345	—0.055
8	*0.943*	—0.606	—0.111
9	0.290	—0.130	*1.000*
10	—0.051	—0.538	*0.980*

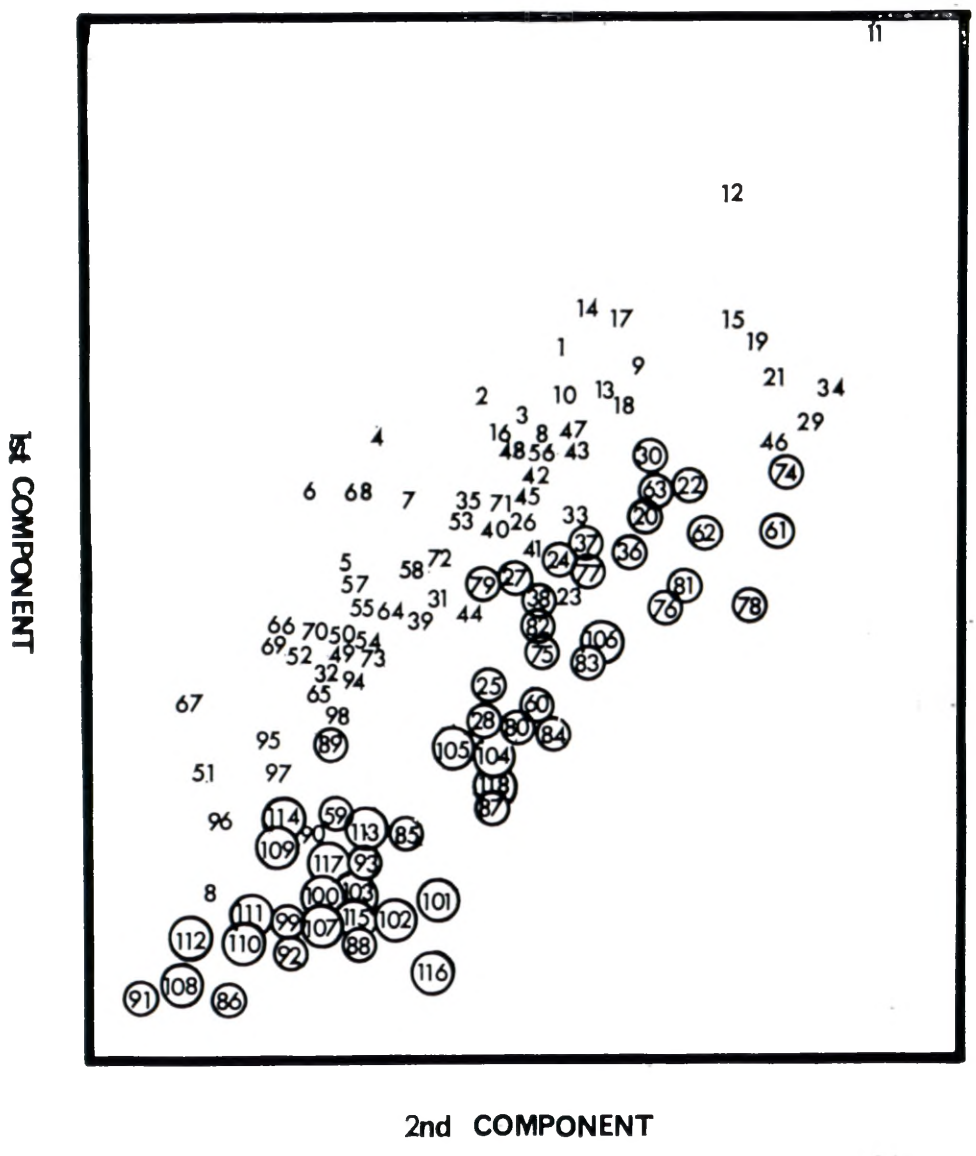

FIG. 7. — Positions of **A. burkei** tree means plotted against the first and second components of the fourth analysis. "Small" populations are circled.

(47 per cent) is extracted by the first component and a total of 81 per cent is extracted by the first three components together. Comparison of Tables 4 and 5 shows only differences of degree within each eigenvector.

The gradients and apparent division discussed under the third analysis are caused by the use of means of means (cf. first and second analyses) as no such discontinuities are obvious from Figure 7 where tree means were used. The "small" populations tend to be clustered along the lower half of the elongated scatter but merge completely with trees from the "big" populations. The scatter in Figure 7 shows that there is a greater difference between trees number 11 and 91 than between trees 6 and 78. Thus there is much more variation within both the "big" and the "small" populations than between them.

Conclusions

A. burkei and *A. nigrescens* are closely related, but nevertheless distinct species, that are readily distinguished from each other on the degree of pubescence of the calyx. In the absence of this taxonomically significant character *A. burkei* and *A. nigrescens* still separate (see Figure 4) although there is no absolute discontinuity. The analysis indicates that rachilla length, number of pinna pairs, number of leaflet pairs, leaflet length and leaflet width are additional characters that enable *A. burkei* and *A. nigrescens* to be differentiated.

There is a rather ill-defined tendency for the appearance of a discontinuity within the *A. burkei* populations. However, contrary to expectations, this discontinuity does not differentiate the "big leaflet" trees from the "small leaflet" trees for both "big" and "small" leaflet plants occur on either side of the discontinuity. Leaflet width is, therefore, not the most important character in creating this discontinuity between the *A. burkei* populations. Petiole length, rachis length, leaf length, number of pinna pairs, number of leaflet pairs, leaflet length and leaflet width are all important characters in creating this discontinuity when considered collectively. Past emphasis on leaflet width alone, a character that provides a rapid visual assessment, as a means of loosely distinguishing between "big leaflet" and "small leaflet" *A. burkei* has tended to obscure the many characters that do contribute to the range of variability within the species.

A. burkei is an extremely variable species and although the specimens at either extreme of the range of morphological variation appear distinctive it is not possible to divide this range of variation satisfactorily and thereby facilitate the recognition of infraspecific categories. As concluded previously (Ross 1968b) it is therefore of doubtful value to recognize infraspecific categories within *A. burkei*.

As a technique for studying taxonomic and ecological problems, principal components analysis is gaining in popularity overseas. After its convincing performance in the present study we hope its popularity will spread to South Africa. One of its attributes is its ability to stimulate further investigation. The reason for the apparently greater variation within *A. burkei* is an example of this stimulus.

Acknowledgements

We would like to thank Dr. D. Edwards and Mr. J. N. R. Jeffers for reading and commenting on this paper prior to its publication.

References

GORDON-GRAY, K. D., 1965. *Acacia robusta* Burch. and *Acacia clavigera* E. Mey. in Natal, South Africa. *Brittonia* 17: 202-213.

JEFFERS, J. N. R., 1965. Principal component analysis in taxonomic research. *Statistics Sect. Pap. For. Commn*, 83.

JESSEN, R J., 1955. Determining the fruit count on a tree by randomized branch sampling. *Biometrics* 11: 99-109.

KENDALL, M. G., 1957. *A course in multivariate analysis.* London: Griffin.

ROSS, J H, 1968a. *Acacia nigrescens* Oliv. in Africa with particular reference to Natal. *Bol. Soc. Brot.*, Sér. 2, 42: 181-205.

ROSS, J. H., 1968b. *Acacia burkei* Benth. in Southern Africa, with particular reference to Natal. *Bol. Soc. Brot.*, Sér. 2, 42: 275-304.

SEAL, H., 1964. *Multivariate statistical analysis for biologists.* London: Methuen.